**Trend Analysis of Recent Twitter #Hashtags**Data Analytics Boot Camp Project 1  
Adam Zimmerman, James Holland, Joshua Gohlike

Proposal: We will be doing an analysis of three recent twitter hashtags to draw conclusions about user demographic and location trends, as well as hashtag performance over time. Specifically:

1. Is there a relationship between user age and the use of hashtags?
2. What can user location tell us about these hashtag trends?
3. How have these hashtags trended in frequency over time?
4. Is there a relationship between hashtag usage and time of day?

We will attempt to answer these questions for the following hashtags:

* #WalkOut
* #Metoo
* #FakeNews

Data source: To answer these questions, we will be exploring twitter APIs to gather the necessary data.

Anticipated tasks:

1. Develop systems for:
   1. productive meetings
   2. effective project management
   3. optimal use of collaborative technology (GitHub, google docs)
2. Explore data sets available through twitter APIs
3. Develop code to collect data through APIs
4. Develop code to explore and clean data
5. Develop code to format and analyze data
6. Graph data and format for meaningful interpretation
7. For presentation, develop a narrative that draws connections between research questions and data analysis and visuals

INTERNAL NOTES:

Did we land on only looking at data in the US? (This will be a filter for data pulls/data cleanup)  
  
Research questions and graphs:

* For age demographics: pie charts or bar graphs
* For location information: heat maps
* For trends over time: line graphs
  + Is there anything to looking at original tweets and retweets as two lines on the graph? Just a thought
* For time of day: ??

Tasks:

* Research appropriate APIs to gather the data necessary to answer our questions - **due date 9/8/18**
  + I believe the Reference API is what we are looking for when looking at location data
* Develop effective processes for successful project completion – **due date 9/8/18**
  + Setting expectations for meetings (agendas, outputs)
  + Divvying up of tasks for completion
  + Task completion expectations
  + Using shared resources/managing merge conflicts (how are we utilizing git branches and resolving conflicts when they occur?)
* Develop project timeline **– due date 9/8/18**
  + Establish due dates and person responsible for tasks

Data frame structure (one data frame per hashtag):

* Rows = tweets
* Columns = data points

Code sections:

* API Requests (one per hashtag, hopefully free account will not limit how much data we can pull)
* Loop through json’s to pull and clean relevant data (break down in loop to improve processing efficiency and tracking)
* Add data to data frames
* Organize and summarize data (groupby, means, sums, etc.)
* Graph formatted data frames
* Do Statistical Analysis (are relationships meaningful?)

Other things to keep in mind:

* Associate specific questions with each visual

Personal interests in project:

Adam:

* Doing loops to fill in the data frames
* Possibly compare to Trumps tweets – look for key words

Jim:

* Place holder

Josh:

* Reinforcing coding that we’ve learned so far
* Pulling data off of the internet, through APIs or otherwise